Generalized Polyhedral Approximations in Convex Optimization

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 17, 6.253 Class, Spring 2014
Lecture Summary

- Outer/inner linearization and their duality.

- A unifying framework for polyhedral approximation methods.

- Includes classical methods:
  - Cutting plane/Outer linearization
  - Simplicial decomposition/Inner linearization

- Includes new methods, and new versions/extensions of old methods.
Extended monotropic programming (EMP)

\[
\min_{(x_1, \ldots, x_m) \in S} \sum_{i=1}^{m} f_i(x_i)
\]

where \( f_i : \mathbb{R}^{n_i} \rightarrow (-\infty, \infty] \) is a convex function and \( S \) is a subspace.

The dual EMP is

\[
\min_{(y_1, \ldots, y_m) \in S^\perp} \sum_{i=1}^{m} f_i^*(y_i)
\]

where \( f_i^* \) is the convex conjugate function of \( f_i \).

Algorithmic Ideas:
- Outer or inner linearization for some of the \( f_i \)
- Refinement of linearization using duality

Features of outer or inner linearization use:
- They are combined in the same algorithm
- Their roles are reversed in the dual problem
- Become two (mathematically equivalent dual) faces of the same coin
Advantage over Classical Cutting Plane Methods

- The refinement process is much faster.
  - Reason: At each iteration we add multiple cutting planes (as many as one per component function $f_i$).
  - By contrast a single cutting plane is added in classical methods.

- The refinement process may be more convenient.
  - For example, when $f_i$ is a scalar function, adding a cutting plane to the polyhedral approximation of $f_i$ can be very simple.
  - By contrast, adding a cutting plane may require solving a complicated optimization process in classical methods.


Outline

1. Polyhedral Approximation
   - Outer and Inner Linearization
   - Cutting Plane and Simplicial Decomposition Methods

2. Extended Monotropic Programming
   - Duality Theory
   - General Approximation Algorithm

3. Special Cases
   - Cutting Plane Methods
   - Simplicial Decomposition for $\min_{x \in C} f(x)$
Given a convex function \( f : \mathbb{R}^n \rightarrow (-\infty, \infty] \).

Approximation using subgradients:

\[
\max \{ f(x_0) + y'_0(x - x_0), \ldots, f(x_k) + y'_k(x - x_k) \}
\]

where

\[
y_i \in \partial f(x_i), \quad i = 0, \ldots, k
\]
Convex Hulls

- Convex hull of a finite set of points $x_i$

- Convex hull of a union of a finite number of rays $R_i$
Inner Linearization - Epigraph Approximation by Convex Hulls

Given a convex function $h : \mathbb{R}^n \mapsto (-\infty, \infty]$ and a finite set of points $y_0, \ldots, y_k \in \text{dom}(h)$

Epigraph approximation by convex hull of rays $\{(y_i, w) | w \geq h(y_i)\}$
Conjugacy of Outer/Inner Linearization

- Given a function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ and its conjugate $f^*$.
- The conjugate of an outer linearization of $f$ is an inner linearization of $f^*$.

Subgradients in outer lin. $\iff$ Break points in inner lin.
Cutting Plane Method for $\min_{x \in C} f(x)$ ($C$ polyhedral)

- Given $y_i \in \partial f(x_i)$ for $i = 0, \ldots, k$, form
  \[ F_k(x) = \max \{ f(x_0) + y'_0(x - x_0), \ldots, f(x_k) + y'_k(x - x_k) \} \]
  and let
  \[ x_{k+1} \in \arg \min_{x \in C} F_k(x) \]

- At each iteration solves LP of large dimension (which is simpler than the original problem).
Simplicial Decomposition for min\(_{x \in C} f(x)\) (f smooth, C polyhedral)

- At the typical iteration we have \(x_k\) and \(X_k = \{x_0, \tilde{x}_1, \ldots, \tilde{x}_k\}\), where \(\tilde{x}_1, \ldots, \tilde{x}_k\) are extreme points of \(C\).
- Solve LP of large dimension: Generate
  \[
  \tilde{x}_{k+1} \in \arg \min_{x \in C} \{(\nabla f(x_k))' (x - x_k)\}
  \]
- Solve NLP of small dimension: Set \(X_{k+1} = \{\tilde{x}_{k+1}\} \cup X_k\), and generate \(x_{k+1}\) as
  \[
  x_{k+1} \in \arg \min_{x \in \text{conv}(X_{k+1})} f(x)
  \]

- Finite convergence if \(C\) is a bounded polyhedron.
Comparison: Cutting Plane - Simplicial Decomposition

- **Cutting plane** aims to use LP with same dimension and smaller number of constraints.

  Most useful when problem has small dimension and:
  
  - There are many linear constraints, or
  - The cost function is nonlinear and linear versions of the problem are much simpler

- **Simplicial decomposition** aims to use NLP over a simplex of small dimension [i.e., \( \text{conv}(X_k) \)].

  Most useful when problem has large dimension and:
  
  - Cost is nonlinear, and
  - Solving linear versions of the (large-dimensional) problem is much simpler (possibly due to decomposition)

  The two methods appear very different, with unclear connection, despite the general conjugacy relation between outer and inner linearization.

  We will see that they are **special cases of two methods that are dual** (and mathematically equivalent) to each other.
Extended Monotropic Programming (EMP)

\[
\min_{(x_1,\ldots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i)
\]

where \( f_i : \mathbb{R}^{n_i} \to (-\infty, \infty] \) is a closed proper convex, \( S \) is subspace.

- Monotropic programming (Rockafellar, Minty), where \( f_i \): scalar functions.
- Single commodity network flow (\( S \): circulation subspace of a graph).
- Block separable problems with linear constraints.
- Fenchel duality framework: Let \( m = 2 \) and \( S = \{ (x, x) \mid x \in \mathbb{R}^n \} \). Then the problem
  \[
  \min_{(x_1,x_2)\in S} f_1(x_1) + f_2(x_2)
  \]
  can be written in the Fenchel format
  \[
  \min_{x\in\mathbb{R}^n} f_1(x) + f_2(x)
  \]

- Conic programs (second order, semidefinite - special case of Fenchel).
- Sum of functions (e.g., machine learning): For \( S = \{ (x, \ldots, x) \mid x \in \mathbb{R}^n \} \)
  \[
  \min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} f_i(x)
  \]
Dual EMP

- Derivation: Introduce \( z_i \in \mathbb{R}^{n_i} \) and convert EMP to an equivalent form

\[
\min_{(x_1, \ldots, x_m) \in S} \sum_{i=1}^{m} f_i(x_i) \quad \equiv \quad \min_{z_i = x_i, \ i=1, \ldots, m, \ (x_1, \ldots, x_m) \in S} \sum_{i=1}^{m} f_i(z_i)
\]

- Assign multiplier \( y_i \in \mathbb{R}^{n_i} \) to constraint \( z_i = x_i \), and form the Lagrangian

\[
L(x, z, y) = \sum_{i=1}^{m} f_i(z_i) + y_i'(x_i - z_i)
\]

where \( y = (y_1, \ldots, y_m) \).

- The dual problem is to maximize the dual function

\[
q(y) = \inf_{(x_1, \ldots, x_m) \in S, \ z_i \in \mathbb{R}^{n_i}} L(x, z, y)
\]

- Exploiting the separability of \( L(x, z, y) \) and changing sign to convert maximization to minimization, we obtain the dual EMP in symmetric form

\[
\min_{(y_1, \ldots, y_m) \in S^\perp} \sum_{i=1}^{m} f_i^*(y_i)
\]

where \( f_i^* \) is the convex conjugate function of \( f_i \).
Optimality Conditions

- There are powerful conditions for strong duality $q^* = f^*$ (generalizing classical monotropic programming results):
  - **Vector Sum Condition for Strong Duality:** Assume that for all feasible $x$, the set
    \[
    S^\perp + \partial \epsilon (f_1 + \cdots + f_m)(x)
    \]
    is closed for all $\epsilon > 0$. Then $q^* = f^*$.
  - **Special Case:** Assume each $f_i$ is finite, or is polyhedral, or is essentially one-dimensional, or is domain one-dimensional. Then $q^* = f^*$.
  - By considering the dual EMP, “finite” may be replaced by “co-finite” in the above statement.

- **Optimality conditions**, assuming $-\infty < q^* = f^* < \infty$:
  - $(x^*, y^*)$ is an optimal primal and dual solution pair if and only if
    \[
    x^* \in S, \quad y^* \in S^\perp, \quad y_i^* \in \partial f_i(x_i^*), \quad i = 1, \ldots, m
    \]
  - Symmetric conditions involving the dual EMP:
    \[
    x^* \in S, \quad y^* \in S^\perp, \quad x_i^* \in \partial f_i^*(y_i^*), \quad i = 1, \ldots, m
    \]
Outer Linearization of a Convex Function: Definition

- Let \( f : \mathbb{R}^n \mapsto (-\infty, \infty] \) be closed proper convex.
- Given a finite set \( Y \subset \text{dom}(f^*) \), we define the outer linearization of \( f \)

\[
f_Y(x) = \max_{y \in Y} \{ f(x_y) + y'(x - x_y) \}
\]

where \( x_y \) is such that \( y \in \partial f(x_y) \).
Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be closed proper convex.
Given a finite set $X \subset \text{dom}(f)$, we define the inner linearization of $f$ as the function $\bar{f}_X$ whose epigraph is the convex hull of the rays $\{(x, w) \mid w \geq f(x), x \in X\}$:

$$\bar{f}_X(z) = \begin{cases} \min \sum_{x \in X} \alpha_x x = z, \sum_{x \in X} \alpha_x = 1, \alpha_x \geq 0, x \in X & \text{if } z \in \text{conv}(X) \\ \infty & \text{otherwise} \end{cases}$$
Polyhedral Approximation Algorithm

Let $f_i : \mathbb{R}^{n_i} \mapsto (-\infty, \infty]$ be closed proper convex, with conjugates $f_i^*$. Consider the EMP

$$\min_{(x_1, \ldots, x_m) \in S} \sum_{i=1}^{m} f_i(x_i)$$

Introduce a fixed partition of the index set:

$$\{1, \ldots, m\} = I \cup \bar{I} \cup \bar{I}, \quad I : \text{Outer indices}, \quad \bar{I} : \text{Inner indices}$$

**Typical Iteration:** We have finite subsets $Y_i \subset \text{dom}(f_i^*)$ for each $i \in I$, and $X_i \subset \text{dom}(f_i)$ for each $i \in \bar{I}$.

Find primal-dual optimal pair $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_m)$, and $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_m)$ of the approximate EMP

$$\min_{(x_1, \ldots, x_m) \in S} \sum_{i \in I} f_i(x_i) + \sum_{i \in \bar{I}} f_i^*(\hat{y}_i(x_i)) + \sum_{i \in \bar{I}} f_i^*(\hat{x}_i)$$

Enlarge $Y_i$ and $X_i$ by differentiation:

- For each $i \in I$, add $\tilde{y}_i$ to $Y_i$ where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$
- For each $i \in \bar{I}$, add $\tilde{x}_i$ to $X_i$ where $\tilde{x}_i \in \partial f_i^*(\hat{y}_i)$. 
Enlargement Step for $i$th Component Function

- **Outer**: For each $i \in I$, add $\tilde{y}_i$ to $Y_i$ where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$.

- **Inner**: For each $i \in \tilde{I}$, add $\tilde{x}_i$ to $X_i$ where $\tilde{x}_i \in \partial f_i^*(\hat{y}_i)$. 

---

**Diagram:**

- **Outer Diagram**: Graph showing the function $f_i(x_i)$ with the slope $\hat{y}_i$. The new slope $\tilde{y}_i$ is indicated with a red line.

- **Inner Diagram**: Graph showing the function $\bar{f}_i, x_i(x_i)$ with the slope $\hat{y}_i$. The new point $\tilde{x}_i$ is marked with a red label.
Mathematically Equivalent Dual Algorithm

Instead of solving the primal approximate EMP

$$\min_{(x_1, \ldots, x_m) \in S} \sum_{i \in I} f_i(x_i) + \sum_{i \notin I} f_{i,Y}(x_i) + \sum_{i \notin \bar{I}} \bar{f}_{i,X}(x_i)$$

we may solve its dual

$$\min_{(y_1, \ldots, y_m) \in S^\perp} \sum_{i \in I} f^*_i(y_i) + \sum_{i \notin I} f^*_{i,Y}(y_i) + \sum_{i \notin \bar{I}} \bar{f}^*_{i,X}(x_i)$$

where $f^*_{i,Y}$ and $\bar{f}^*_{i,X}$ are the conjugates of $f_{i,Y}$ and $\bar{f}_{i,X}$.

Note that $f^*_{i,Y}$ is an inner linearization, and $\bar{f}^*_{i,X}$ is an outer linearization (roles of inner/outer have been reversed).

The choice of primal or dual is a matter of computational convenience, but does not affect the primal-dual sequences produced.
In some cases we may use an algorithm that solves simultaneously the primal and the dual.

- Example: Monotropic programming, where $x_i$ is one-dimensional.
- Special case: Convex separable network flow, where $x_i$ is the one-dimensional flow of a directed arc of a graph, $S$ is the circulation subspace of the graph.

In other cases, it may be preferable to focus on solution of either the primal or the dual approximate EMP.

After solving the primal, the refinement of the approximation ($\tilde{y}_i$ for $i \in I$, and $\tilde{x}_i$ for $i \in \bar{I}$) may be found later by differentiation and/or some special procedure/optimization.

- This may be easy, e.g., in the cutting plane method, or
- This may be nontrivial, e.g., in the simplicial decomposition method.

Subgradient duality $[y \in \partial f(x) \iff x \in \partial f^*(y)]$ may be useful.
Cutting Plane Method for $\min_{x \in C} f(x)$

- EMP equivalent: $\min_{x_1=x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of $C$.

- Classical cutting plane algorithm: Outer linearize $f$ only, and solve the primal approximate EMP. It has the form

$$\min_{x \in C} f^*_Y(x)$$

where $Y$ is the set of subgradients of $f$ obtained so far. If $\hat{x}$ is the solution, add to $Y$ a subgradient $\hat{y} \in \partial f(\hat{x})$. 

![Diagram showing the cutting plane method](image)
Simplicial Decomposition Method for $\min_{x \in C} f(x)$

- **EMP equivalent**: $\min_{x_1 = x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of $C$.

- **Generalized Simplicial Decomposition**: Inner linearize $C$ only, and solve the primal approximate EMP. In has the form

$$\min_{x \in \bar{C}_X} f(x)$$

where $\bar{C}_X$ is an inner approximation to $C$.

- Assume that $\hat{x}$ is the solution of the approximate EMP.
  - Dual approximate EMP solutions:
    $$\{(\hat{y}, -\hat{y}) \mid \hat{y} \in \partial f(\hat{x}), -\hat{y} \in \text{(normal cone of } \bar{C}_X \text{ at } \hat{x})\}$$
  - In the **classical case** where $f$ is differentiable, $\hat{y} = \nabla f(\hat{x})$.
  - Add to $X$ a point $\tilde{x}$ such that $-\hat{y} \in \partial \delta(\tilde{x} \mid C)$, or
    $$\tilde{x} \in \arg \min_{x \in C} \hat{y}' x$$
Illustration of Simplicial Decomposition for $\min_{x \in C} f(x)$

Differentiable $f$

Nondifferentiable $f$
Dual Views for $\min_{x \in \mathbb{R}^n} \left\{ f_1(x) + f_2(x) \right\}$

- **Inner linearize $f_2$**

- **Dual view: Outer linearize $f_2^*$**
Assume that
- All outer linearized functions $f_i$ are finite polyhedral
- All inner linearized functions $f_i$ are co-finite polyhedral
- The vectors $\tilde{y}_i$ and $\tilde{x}_i$ added to the polyhedral approximations are elements of the finite representations of the corresponding $f_i$

Finite convergence: The algorithm terminates with an optimal primal-dual pair.

Proof sketch: At each iteration two possibilities:
- Either $(\hat{x}, \hat{y})$ is an optimal primal-dual pair for the original problem
- Or the approximation of one of the $f_i, i \in I \cup \bar{I}$, will be refined/improved

By assumption there can be only a finite number of refinements.
Convergence - Nonpolyhedral Case

- **Convergence, pure outer linearization** ($\bar{I}$: Empty). Assume that the sequence $\{\tilde{y}^k_i\}$ is bounded for every $i \in \bar{I}$. Then every limit point of $\{\hat{x}^k\}$ is primal optimal.

- **Proof sketch**: For all $k, \ell \leq k - 1$, and $x \in S$, we have

\[
\sum_{i \notin \bar{I}} f_i(\hat{x}^k_i) + \sum_{i \in \bar{I}} \left( f_i(\hat{x}^\ell_i) + (\hat{x}^k_i - \hat{x}^\ell_i)'\tilde{y}^\ell_i \right) \leq \sum_{i \notin \bar{I}} f_i(\hat{x}^k_i) + \sum_{i \in \bar{I}} f_{i,Y_{i-1}^k}(\hat{x}^k_i) \leq \sum_{i=1}^m f_i(x_i)
\]

- Let $\{\hat{x}^k\}_{\mathcal{K}} \rightarrow \bar{x}$ and take limit as $\ell \rightarrow \infty$, $k \in \mathcal{K}, \ell \in \mathcal{K}, \ell < k$.

- Exchanging roles of primal and dual, we obtain a convergence result for pure inner linearization case.

- **Convergence, pure inner linearization** ($\bar{I}$: Empty). Assume that the sequence $\{\tilde{x}^k_i\}$ is bounded for every $i \in \bar{I}$. Then every limit point of $\{\hat{y}^k\}$ is dual optimal.

- **General mixed case**: Convergence proof is more complicated (see the Bertsekas and Yu paper).
Concluding Remarks

- A unifying framework for polyhedral approximations based on EMP.
- Dual and symmetric roles for outer and inner approximations.
- There is option to solve the approximation using a primal method or a dual mathematical equivalent - whichever is more convenient/efficient.
- Several classical methods and some new methods are special cases.
- Proximal/bundle-like versions:
  - Convex proximal terms can be easily incorporated for stabilization and for improvement of rate of convergence.
  - Outer/inner approximations can be carried from one proximal iteration to the next.